ABSTRACT

What is the future of cataloging in light of ongoing developments in digital technologies? At one extreme are those who believe it is up to current institutions to take on the task of cataloging, and generally organizing, digital materials; at the other are those who imagine new sorts of tools and techniques will obviate the need for these traditional methods. This paper, written by a computer scientist, explores the nature of cataloging as it is now practiced. Drawing on textbooks and other library science materials, as well as recent results in anthropology and sociology, it characterizes cataloging as a form of order-making, invisibly sustaining "the order of the book." It then looks at how this order is being challenged by changes tied to the wide-scale adoption of digital technologies, and concludes by exploring the cataloging of digital materials.

KEYWORDS: Cataloging, Digital documents, Digital libraries, Internet, World-Wide Web.

INTRODUCTION

The explosive growth of the Internet and the World-Wide Web[1] in recent times has caused great excitement in many quarters. Suddenly an extraordinary wealth of materials on a huge range of subjects is instantaneously available from desktops all over the world. This excitement is tempered somewhat by the awareness that this material is poorly organized, and of variable quality and stability; it is difficult to conceptualize, browse, search, filter, or reference. Increasingly, people are wondering how this situation can be improved. At one extreme are those who believe it is up to current institutions to take on the task of cataloging, and generally organizing, these materials; at the other are those who imagine new sorts of tools and techniques -- sophisticated search tools and "agents" -- will obviate the need for these traditional methods.

During the past year, roughly since Digital Libraries '94, I too have begun to reflect on these issues. Trained as a computer scientist, I already had a reasonable grasp of the computational state of the art, but was lacking in any depth of knowledge of relevant library practices. And so, with the help of a number of people -- professors and students of library science as well as practicing librarians (see acknowledgments below) -- I set out to gain some understanding of cataloging practices. During this time I have consulted and read textbooks and articles on cataloging, talked with catalogers[2], and, in one case, observed a cataloger at work. I have also been a subscriber, and a regular reader, of email on various library/cataloging listservs, including Autocat, Pacs-I, Emedia, Web4Lib, and Intercat. The first of these lists in particular has given me considerable insight into the kinds of questions and concerns that catalogers voice among themselves -- those they are willing to verbalize over the Net, at any rate.

My aim in this paper is to articulate some of the understanding I have gleaned from these activities, and to apply it to questions of cataloging digital materials. What is cataloging, anyway? Why has it been important, and to what extent will it remain so in the future? It is always a tricky business to venture out beyond one's training and to attempt to characterize the work of others, but I believe that there is merit in this, so long as one realizes the limits of such endeavors, and stands ready to be corrected by those more knowledgeable. Besides, this seems to be a time when all of us must venture, somewhat uneasily, outside our primary areas of expertise and engage in dialog over issues of mutual concern.

I will proceed as follows: In the next section I will briefly summarize my understanding of library cataloging, as articulated in textbooks and other teaching materials. Following this, drawing on recent research in the anthropology of work and the history of the book, I will frame cataloging as a kind of ongoing order-making, invisibly sustaining "the order of the book." I will then look at how this order is being challenged by changes tied to the wide-scale adoption of digital technologies, and I will conclude by considering the cataloging of digital
WHAT IS CATALOGING?

Library cataloging\[3\] is the practice of organizing a collection of bibliographic items to facilitate their identification, location, access, and use. Perhaps the simplest way to do this is to organize the items themselves, creating what Miksa calls an "item file" -- e.g. by arranging papers in a filing cabinet or books on a shelf. This method works well for small collections, but for large collections more elaborate means are needed.

The principal method for organizing larger collections is to develop a catalog. A catalog consists of a set of entries, each of which stands for an item in the collection and which describes certain characteristics of the item, such as (for a book) its author, title, publisher, subjects, and so on. The catalog is itself a collection -- a collection of surrogates for items in the primary collection; these surrogates must be arranged as well. There is a highly articulated set of strategies for organizing catalogs, e.g. alphabetically by author and/or by subject.

Current cataloging practices involve both strategies. Surrogates are created for the items and arranged in a catalog. The items themselves are also arranged -- e.g. books are placed "linearly" on the shelves of a library's stacks. This is typically accomplished via a classification scheme, such as the Library of Congress Classification (LCC) or the Dewey Decimal Classification (DDC), in which a hierarchy of possible subjects is given a linear ordering and each subject is given its place on the shelves. This means that an item's call number not only specifies an item's location in the stacks but collocates it with other items that address the same putative topic.

A distinction is made between two types of cataloging activity, both of which are practiced to catalog a particular item: descriptive cataloging and subject cataloging. Descriptive cataloging is concerned with creating catalog records for items, describing their characteristics as just noted -- author, title, and so on. Subject cataloging is concerned with classifying the subject matter, the intellectual content, of an item. It is the subject cataloger who assigns an item to a class within a classification scheme which in turn determines a place on the shelf.

A distinction is also made between bibliographies and catalogs. Both of these in practice describe items. The difference is that a bibliography describes works and editions of works, but not actual physical items. A catalog, by contrast, primarily describes particular, physical items in a particular collection. It does this partly by describing aspects of a work (e.g. title and author), as does a bibliography, and partly by indicating physical properties, including its location. The bibliography at the end of this paper, for example, lists the seventh edition of Wynar's *Introduction to Cataloging and Classification* but makes no reference to the physical copy I checked out of Stanford Library. The catalog record I consulted at Stanford gave me the call number of a particular copy of the seventh edition (Z693 W94 1985), the only copy in Stanford's collection.

The development of highly sophisticated, systematically organized catalogs and cataloging procedures is a product of the modern library era, which dates from the second half of the last century. Book catalogs (lists of entries bound in book form) were the first kind used in the U.S.; these began to be displaced by the familiar card catalogs around the turn of the century. Digital catalogs (called OPACs, Online Public Access Catalogs) began to appear in the 1970's and are now widespread; they are rapidly displacing card catalogs. Entries in OPACs are commonly encoded in MARC (Machine Readable Cataloging) format, a standard which permits them to be shared among institutions. To a great extent, the MARC standard encodes those features previously recorded on cards; it enables a fairly straightforward translation of card contents into digital form.

In libraries today, cataloging is considered part of "technical services" -- those services generally concerned with the maintenance of the collection, such as acquisition and binding. Technical services are distinguished from "public services," such as reference services, which involve direct contact with library users. Nearly all libraries have catalogers, although in very small libraries one person may handle other tasks, e.g. reference work and acquisition of new materials, in addition to cataloging.

The cataloger's job is to produce catalog records for newly acquired materials. "Original cataloging" is the process of creating catalog records from scratch -- creating a record primarily using the item itself, e.g. by inspection of the book, including, but not limited to, its title page. This is distinguished from "copy cataloging," in which the cataloger makes use of a previously existing catalog record for the item to create a new record tailored to the needs of their own library. Catalog records that can be used as sources for copy cataloging are maintained by several institutions, including the Library of Congress, OCLC, and RLG. While it may take only a few minutes to copy catalog an item, it is not unusual for the original cataloging of an item to take on the order of an hour.

CATALOGING AS ORDER-MAKING

It should be clear from this description that cataloging is a form of order-making; it is a set of practices which quite literally put a library's collections in order and provide access through a set of systematically organized surrogates; and it is therefore a crucial part of the system by which books and other materials are maintained and made available to readers throughout the world. The work is ever ongoing, not just because new items are constantly being acquired, but because a collection will crumble -- like the old barns one sees in the countryside -- unless it is continually maintained: books disappear and need to be reordered, they are damaged and need to be repaired, they are misshelved, and so on; catalogs need to be similarly maintained. This might seem like an obvious enough point, but cataloging shares with many other forms of order-making a certain invisibility 20] For the ordinary library patron, the machinery by which the collection is maintained is out-of-sight, ill-understood, and the order it produces taken for granted.

The typical library patron might be excused for not knowing how much work is being done behind the scenes to make an item available;
But to many of us in the [library] profession, catalogers remain an unfathomable mystery. Unable to understand the fires of dedication that burn brightly within them, we often poke fun at their internecine controversies. To those of us who are on the firing line of big issues like intellectual freedom and library funding, the wars waged by catalogers over the future of the main entry or the role of the hyphen often appear to be peevish squabbles fought by socially dysfunctional nitpickers. Where did catalogers come from we often wonder. Theories, of course, abound. Some speculate that they are aliens from a faraway galaxy who have come to earth to tidy things up a bit. Others believe that catalogers may be the descendants of the lost tribe of Israel. After all, they point out, there is a very close similarity between the book of Deuteronomy and AACR2.

Although these remarks are probably somewhat tongue-in-cheek, it is striking how similar they are to assessments made of other forms of service work -- such as office work and machine repair -- which are often taken to be procedural and to require little skill or intelligence, and which are often performed by women or other under-recognized groups. Is the work of cataloging simply routine, a matter of filling in blank catalog records by transcribing certain features from the book or journal? (What is the title? Simply look at the title page. Who is the author? Ditto.) Could this be done by chimpanzees, or by computers?

The answer, I believe, is no: cataloging is not simply a matter of reading off self-evident properties of items, but is a highly skilled interpretative activity by which the properties of items are not simply described, but stabilized and even created. My main evidence for this view comes from reading the mail on the Autocat listserv. Autocat is a heavily used discussion group for catalogers from around the world (with on average 30 to 50 messages per day) and I have found the quality of discussion to be unusually high. Listening in ("lurking") on Autocat has given me an opportunity to attend to aspects of the actual practice of catalogers that could only be surpassed by doing an ethnographic investigation of cataloging, or becoming a cataloger myself.

What has made Autocat particularly valuable to me is that, because of the nature of the medium, list members must articulate their views and questions, so that aspects of what is normally tacit practice and belief are made "publicly" available. Over the last year, discussion has ranged over many topics, including: how and whether to catalog digital documents, the future of cataloging, the politics of cataloging, the relationship between cataloging and reference services, teaching cataloging in library schools, and cataloging humor. But perhaps the central core of discussion is a steady, daily stream of questions, in which catalogers ask one another for help with items they are cataloging, for help in interpreting the cataloging rules, and for advice with materials that the cataloging rules don't address.

What emerges is a much more complex -- and interesting -- picture than Manley's parody, or lay understanding, will admit. Bibliographic materials, it appears, do not always wear their bibliographic properties on their sleeve, and the rules for determining and "transcribing" these properties are quite complex and require interpretation to be applied. Anyone who has written a dissertation or a paper requiring a formal bibliography has encountered this problem in a small way: an item that is hard to categorize (is it a serial or a monograph?), that lacks an expected property (who is the author?), or one of whose properties is ambiguous (who is the publisher for a reprint?).

What we were never told in high school or college, when we first learned to write a research paper, is that bibliographic descriptions are idealizations of, or approximations to, the materials they describe. (Of course this is true of all descriptions.) Cataloging is not just a matter of "reading off" the properties items have but of normalizing or regularizing the material to conform to standard categories of description and thereby making the properties in the act of describing them. That books have definitive titles and authors, that volumes have unambiguously of certain types, is a good first approximation to the truth, and the work of cataloging is to maintain this first approximation as a useful and usable fiction. Cataloging appears to be routine work so long as one believes that the materials are unambiguously of certain types, is a good first approximation to the truth, and the work of cataloging is to maintain this first approximation as a useful and usable fiction. Cataloging appears to be routine work so long as one believes that the materials just have a regular structure which can be trivially read off. But on inspection, it appears that this regular structure is the output of the work of catalogers, not the input.

But where did the standard categories of description come from in the first place in terms of which library materials are regularized? They are part and parcel of what the historian of the book, Roger Chartier, has called "the order of books" an order developed “between the end of the Middle Ages and the eighteenth century [in the] attempt to master the enormously increased number of texts that first the manuscript book and then print put into circulation” (p. vii). This order includes a set of complexly interconnected institutions for the production, distribution, conservation, and use of bibliographic materials (publishers, booksellers, libraries, a reading public, etc.); a rich set of constructs for classifying these materials (notions of work, author, edition, etc.); and a regulatory mechanism based on copyright for determining identity and ownership.

Recent scholarship, much of it drawing inspiration from Foucault's article, "What is an Author?" has shown that the notion of authorship "far from being timeless and universal, is a relatively recent formation -- the result of a quite radical reconceptualization of the creative process that culminated less than 200 years ago in the heroic self-presentation of Romantic poets." Moreover, this notion of authorship is intimately connected to other components of the order, most notably to the system of copyright, which by mediating disputes serves to regulate ownership and even identity of works.

Thus the apparently effortless and routine reading off of properties such as authorship is the result of a complex system of ordering practices of which cataloging is an intimate part. The larger order, the order of books, determines the categories by which items can be classified, and catalogers do the invisible work of upholding this order, regularizing and standardizing, nipping and tucking, making it appear that items really do have the "natural" properties we take for granted.

A CHANGING ORDER
None of this is static, of course, it is all in motion, subject to continual maintenance and change. Not only must collections be continually maintained, as noted earlier, but changes in the larger sociotechnical environment present challenges whose effects ripple through the entire order. In this century, the development of film, video, microfilm, and photocopying technologies have all required changes and accommodations in the law, in library practices, in publishing and distribution, in patterns of consumption and use.

None of these developments, however, has produced the uncertainty, confusion, and anxiety that has been occasioned by the increasing adoption of digital technologies, for none of these earlier developments challenged the order of the book, the order built around one central type of physical artifact and its modes of production, conservation, and consumption. But now, virtually every aspect of this order is being questioned: How must publishing change when "publications" can be instantaneously distributed on the Net? What models of compensation are appropriate to these new modes of production and distribution? Whose interests must be given priority? What happens to the notion of edition -- a set of "identical" artifacts produced by a publisher -- when one-of-a-kind, customized documents can be produced on a large scale? What happens to the notion of author, and the distinction between author and reader, if hypertext documents become, as some suggest [3], fluidly modified, collaborative efforts? How must copyright change to accommodate these other shifts, or should it be replaced by a different regulatory system?

Perhaps the greatest uncertainty, from the perspective of cataloging, is just what the new digital materials will be. Our current order carves up the bibliographic universe into (relatively) discrete, stable, and long-lived units. But now, there is the potential at least for a great deal more variability and mutability of materials[6], and for a less rigid boundary between items [13]. New genres, new categories of description, new institutions and practices have not yet arisen to stabilize this material. All of this together would constitute a new order, or substantial changes to the old order, as yet unrealized. (I will call this the "digital order," for lack of a better name, with the understanding that it is certain to include many elements of the order of books.)

Currently there is no lack of visionaries willing to sing the praises of this new order, and to proclaim their vision of the future as good, right, and inevitable. Such confidence seems misguided; I am not at all sure that the future can be extrapolated, or read off, from current trends -- least of all from current technological trends. Carla Hesse argues, for example, that the modern order of the book was created in post-revolutionary France by direct political intervention [10]. From this she draws the conclusion that one of the main determinants of the digital order will be the regulatory framework now being debated in Washington. If she is right, then the next stages will be the product of conscious political choices made in relation to the evolving technologies, not a "natural" and inexorable result of technology development.

CATALOGING IN THE DIGITAL ORDER

Finally, having examined aspects of cataloging and the larger order of which it is a part, I will briefly explore the question of cataloging materials on the Net. Notice that I have said "cataloging materials on the Net" and not "cataloging the Net." From a cataloging perspective, the Net is the wrong unit of analysis, for the Net does not constitute a library in any interesting sense. It may be part of the technological infrastructure which could support a library, but it is not a library -- not in the sense of an institution which oversees and provides access to a collection. Moreover, the materials on the Net do not constitute a library collection -- not in the sense of a selection of items organized to serve a particular clientele [14]. Instead, we should probably talk about cataloging digital materials on the Net. (In this same spirit, OCLC refers to "cataloging Internet resources" rather than "cataloging the Internet." [17])

There is of course a tremendous amount of activity in this area at the moment:

- Relatively large scale catalogs of and indices into Internet resources have been created and are being extended. Yahoo[9] and Alex[10], each with thousands of entries, are but two of the many catalogs now coming into existence.
- There are ongoing experiments and feasibility studies in cataloging digital documents, such as the Catriona[11] project funded by the British Library and a project organized by OCLC, called "Building a Catalog of Internet Resources."[12]
- New standards are being created, and existing ones modified, to represent digital documents and "meta-data" about digital documents. The Text Encoding Initiative (TEI) [19], for example, has created a standard, referred to as TEI P3, for the digital encoding of texts in the humanities, which includes a proposal for an electronic text "header" for documents, permitting the equivalent of title page information to be represented, as well as other forms of meta-data. The Working Group on Document Identifiers within the Internet Engineering Task Force (IETF) is working to standardize Uniform Resource Characteristics (URCs) to represent meta-data. And a new field (856) has been added to MARC records so that catalog entries can record the location (e.g. the URL) of digital documents.
- A great deal of information about approaches to cataloging digital documents is appearing on the Net. The Cataloging Services Department of the University of Virginia Library, for example, has a Web page[15] with pointers to many cataloging related resources. There is also lots of discussion of cataloging issues on many Internet lists, including Autocat, Pacs-l, Emedia, Web4Lib, and Intercat.
- Individuals, groups and institutions are creating their own home pages. These often contain lists of "favorite" Net resources that visitors to their home page may want to view. Using Web browsers such as Mosaic, people are also creating and maintaining "hotlists" of Net resources to which they can have easy access.

With all this activity, much is up in the air, much is being questioned and debated. Can anything be said definitively about cataloging in the digital order? Clearly, digital collections will need to be organized and maintained if they are to remain stable and usable. To those within the library community this point may seem too obvious to make, since it is one of the central premises on which the modern
library movement has been built. Yet it is not widely understood outside this community, to judge from the way the Web is often talked about. There seems to be a fairly wide-spread presumption, tacitly held, that materials can maintain their identity and their currency without continual intervention, without continual work.

Moreover, it seems certain that the work of stabilizing and maintaining digital collections will require a great deal of systematic human activity. Certainly, powerful technologies will be involved as well as human labor, as is the case now. Again, this is obvious enough to librarians; technologists, by contrast, tend to see the technical infrastructure but not the "invisible" social infrastructure by which most things, not just library collections, are maintained.

But is there a future specifically for catalogs and cataloging? Is it possible that increasingly powerful search and browsing tools, automated indexing tools, and "intelligent agents" will obviate the need for catalogs? I think not: even if automated tools someday allow people to locate documents without mediation by catalog entries, catalogs will still exist, I am quite sure -- although some care must be taken in using the word "catalog" in this context.

There can be no doubt that people will continue to make lists of digital resources which they will want to make available to others. Some of these will have broad scope and utility (e.g. Yahoo and Alex), while others (personal hotlists and favorite places recorded in home pages) will be of limited scope. Whether we call these lists "bibliographies" or "catalogs," there will be an ongoing need to maintain them, to keep them current and stable; and whether or not we call this ongoing maintenance work "cataloging," it will no doubt be supported by a complex, and largely invisible, sociotechnical infrastructure.

Of course to say that there will continue to be catalogs still leaves open many important questions. What sorts of materials will be cataloged? How will digital catalogs differ from those we have today? Will they be maintained by professional catalogers or by laypeople? What skills will be needed to create these new catalogs? How will such skills be acquired? To what extent will these catalogs be based on existing standards, such as MARC and AACR? Will there, or could there be, a universal catalog of digital materials?

There can be no answers to most of these questions at this time; they will emerge in relation to, and will partly constitute, a stable digital order. Some, however, may already be amenable to analysis and may yield at least partial results, but I think they will require us to bring knowledge from multiple work communities, including those I have juxtaposed in this paper: computer science, library science, history, and anthropology. Take the question of a universal catalog of digital materials. While there is something quite appealing about this vision, research in anthropology [9] suggests that there is an inevitable tension between the desire to generalize, to make universal, and the need to tailor to local conditions. The current use of OCLC and RLIN displays both tendencies: these utilities provide sharable catalog records (the general) which are then copied and modified for local use by member institutions (the local, the specific). What reason is there to think that the desire to make local copies with local annotations will be any the less in the future?

In the same passage from which I quoted earlier, Chartier [5] says:

[How did people in Western Europe between the end of the Middle Ages and the eighteenth century attempt to master the enormously increased number of texts that first the manuscript book and then print put into circulation? Inventorying titles, classifying works, and attributing texts were all operations that made it possible to set the world of the written word in order. Our own age is the direct heir of this immense effort motivated by anxiety. (p. vii)]

If we are indeed, as it now seems, entering into another immense effort of order-making, this time we can benefit from the efforts of our forebears, as well as our own recent scholarship, as we begin to engage in a dialog across communities and disciplines.

ACKNOWLEDGMENTS

Many people have been generous with their time and guided me in my exploration of library issues, including Michael Buckland, Mark Butler, Fran Miksa, Sally Peters, Daniel Pitti, Vicky Reich, Vitus Tang, and Alice Wilder-Hall. For feedback on this paper, I would like to thank Philip Doty, Judith Hopkins, Ted Metcalfe, Sally Peters, Brian Schottlaender, Alice Wilder-Hall, and those who anonymously reviewed the paper for this conference.

REFERENCES

11. Horowitz, L.R., CETH Workshop on Documenting Electronic Texts. 1994, New Brunswick, NJ: Center for Electronic Texts in the

Digital Cataloging Related Article. Levy, D. M. (1995). Cataloguing in the digital order: Paper regarding the future of cataloguing, from the Digital Libraries 95 conference. This article is an overview of digital age cataloging. It begins with an introduction to the general idea of cataloging. The next section of the article, entitled "Cataloging as order-making" describes just that. A discussion of how materials are maintained and, most importantly, made available. It also includes a more in-depth discussion of what cataloging can truly involve, such as discussions of standards. The remainder of the article delves into how cataloging is being affected and changed by the new digital word. PDF | This article expands the concept of original order and presents two systematic views of the order of electronic records. In one view, abstract conceptual order can be used as guidance for the physical arrangement of records storage media and intellectual arrangement of... The traditional MARC (Machine Readable Cataloging) format falls short in representing these hierarchical relationships. Based on detailed analyses of these [Show full abstract].
These catalogues are normally sent on order and have a digital alternative with an identical layout. This means savings and the ability to collect information on interested customers, even before they make a purchase, which will be used in retargeting, although it can also inspire laziness or suspicion for those who don't want to give their details to the shop. To receive their catalogue in the letterbox has become an annual ritual, besides what used to be a false privilege of receiving the publication free of charge, even though there was a price on the spine of the publication. However, the digital version and the app have been a failure due to their functionality problems and the inability to add brands or annotations.